Streaming Data Science and Data Engineering

An example from Mobility

Outline

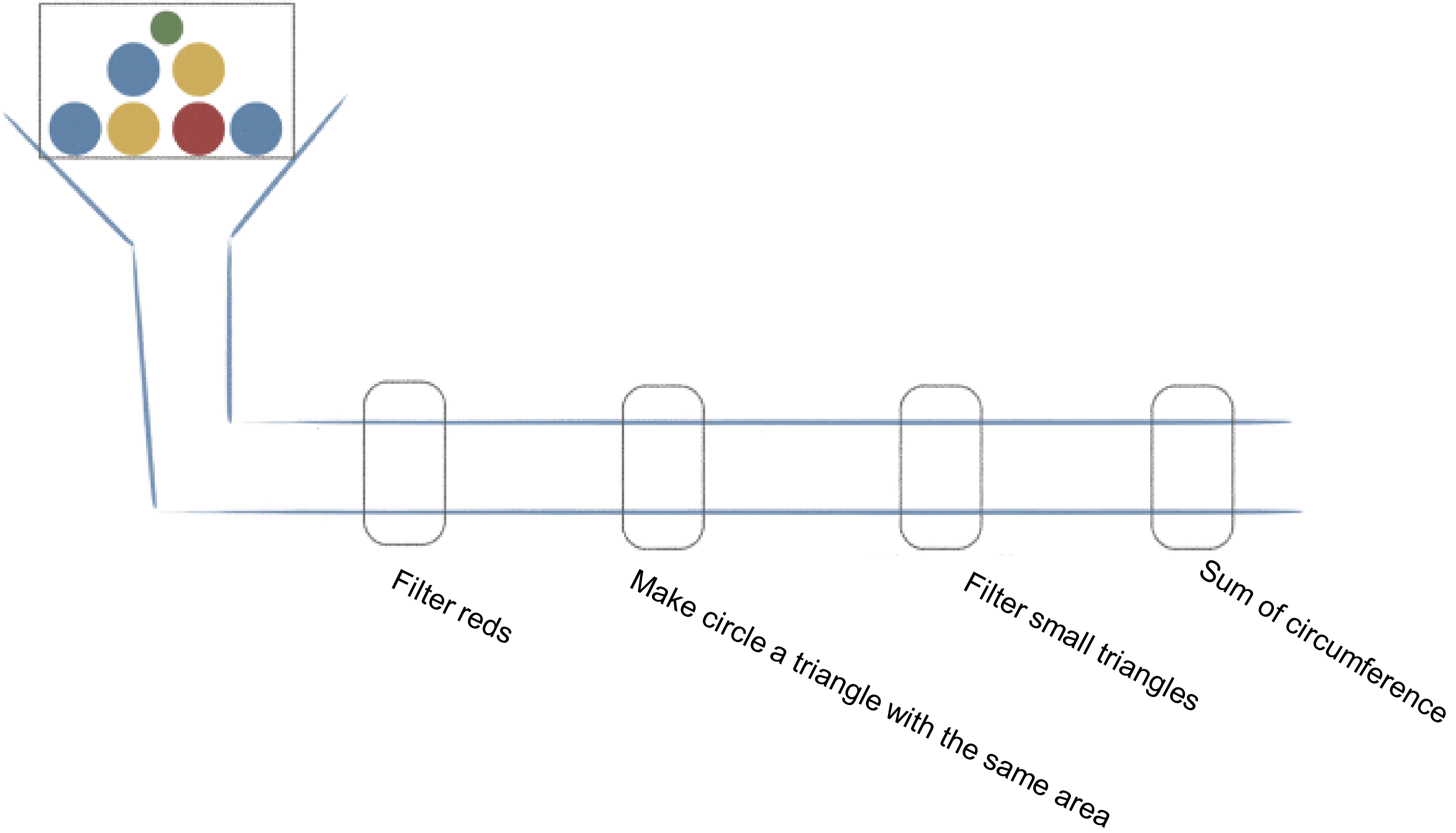
- Streaming Data Science
    - *Concepts & Functions*

- A concrete example: Road Traffic Monitoring and Analysis
    - *Traffic Congestion and Economic Context*
    - *Applied concepts*

- Conclusions

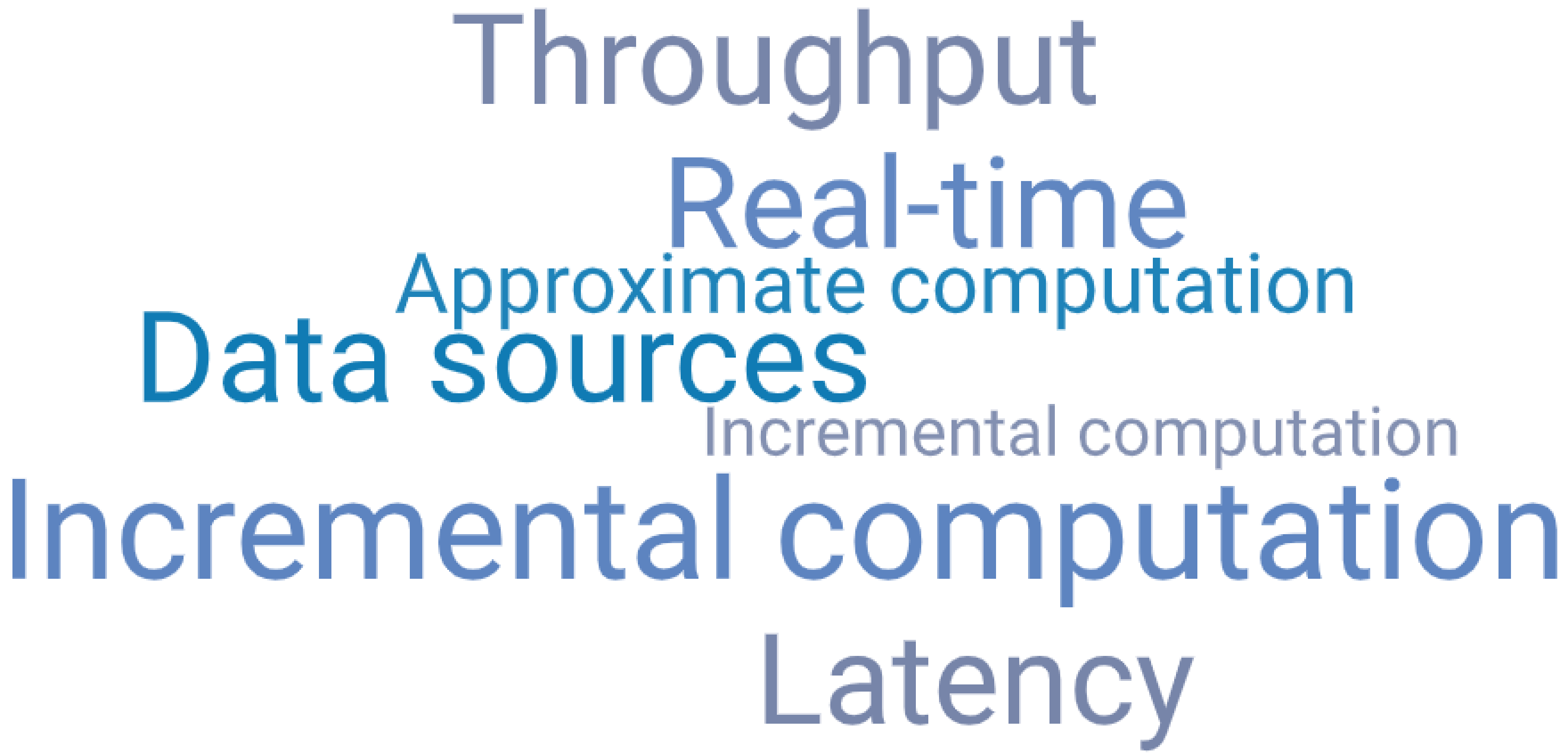
Streaming Data Science

An intuitive introduction

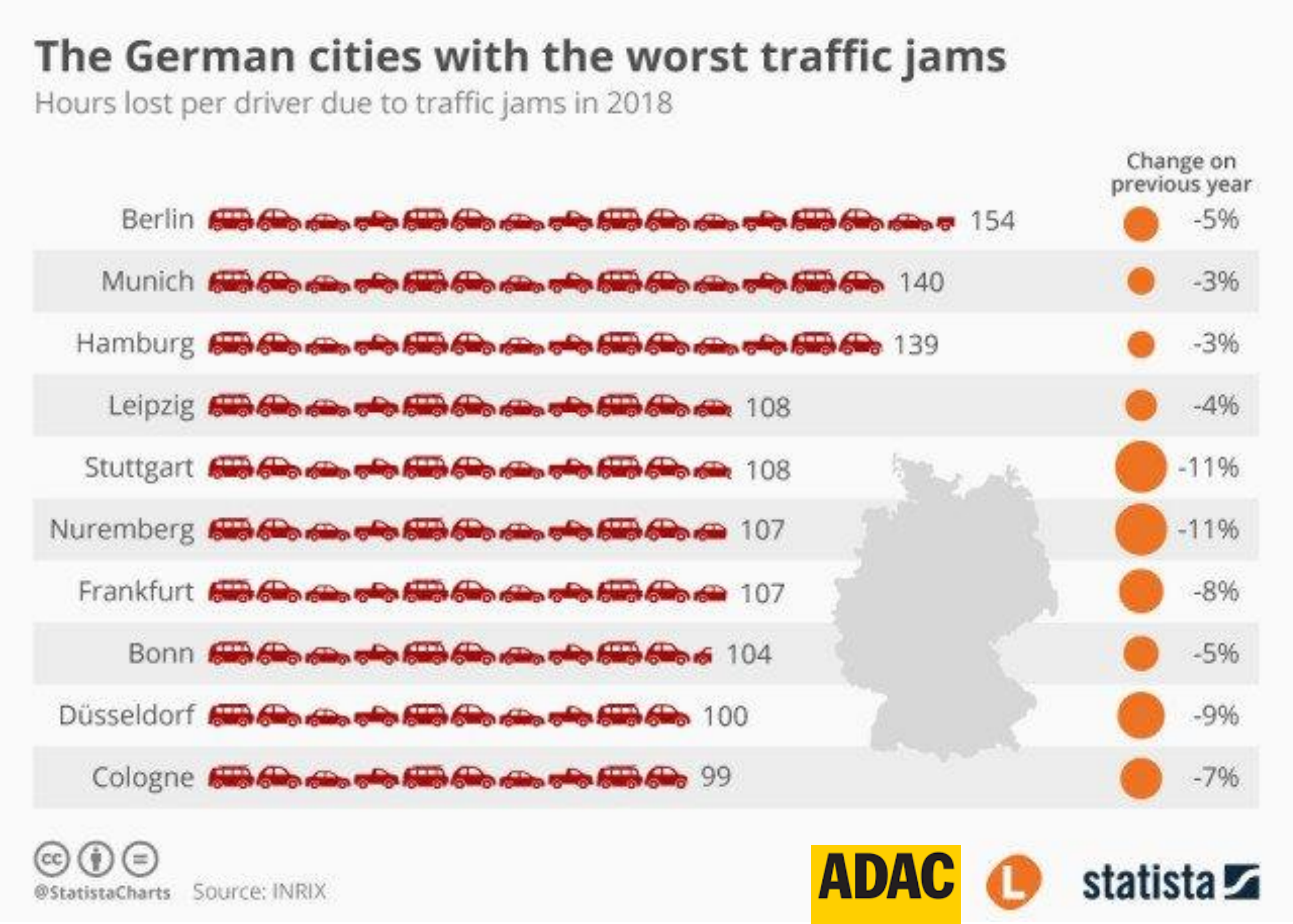


Streaming Data Science

Concepts and functions



Traffic Congestion and Economic Context



Year 2018 in Berlin

- **Total of 1.5 million Km jam**
- **Total of 154 h time lost/driver**

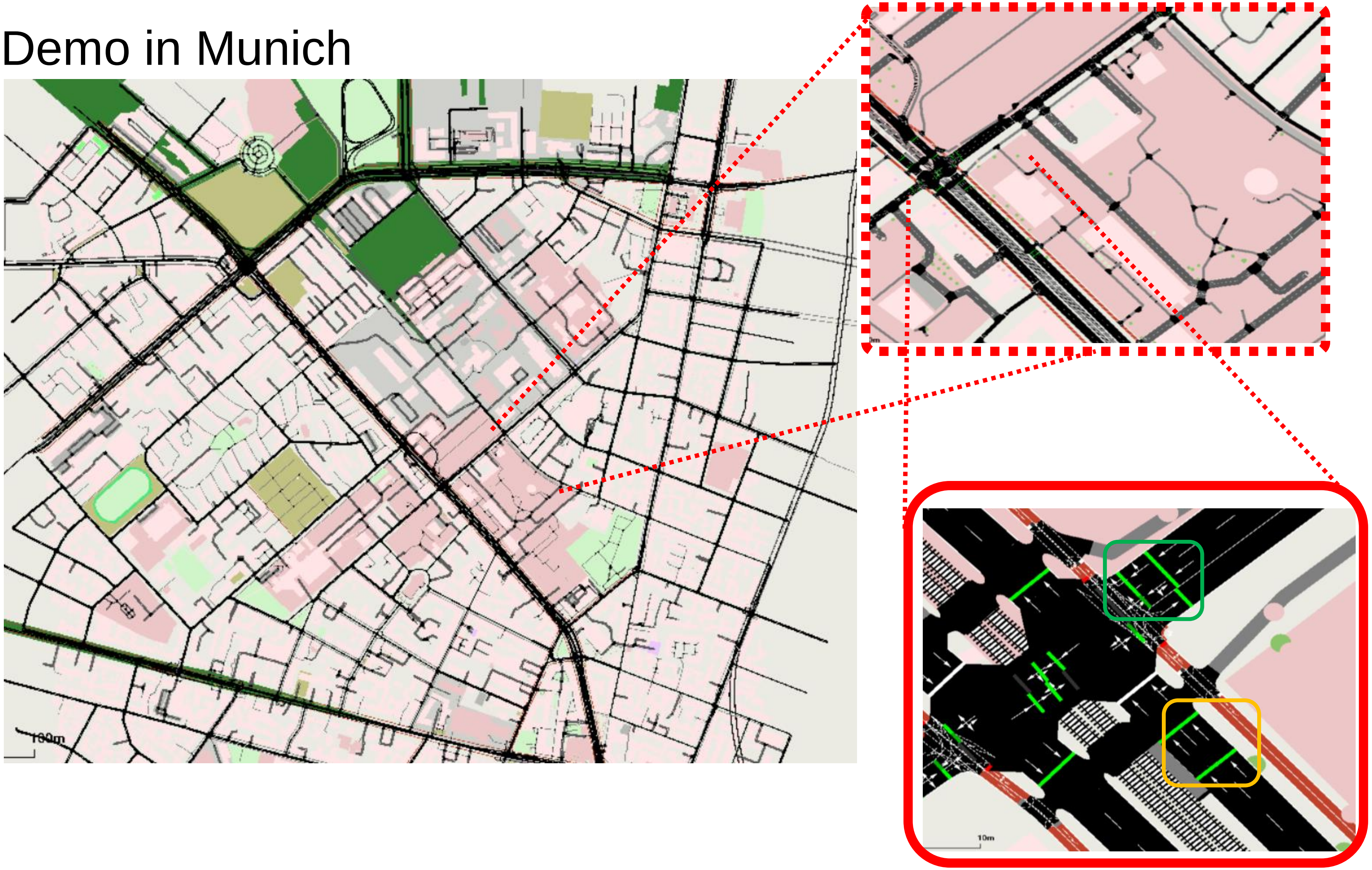
Year 2018 in Germany

- **Average 120 h time lost/driver**
- **Economic loss 1,052 €/driver**

Source: https://www.thelocal.de/20190213/traffic-jams/

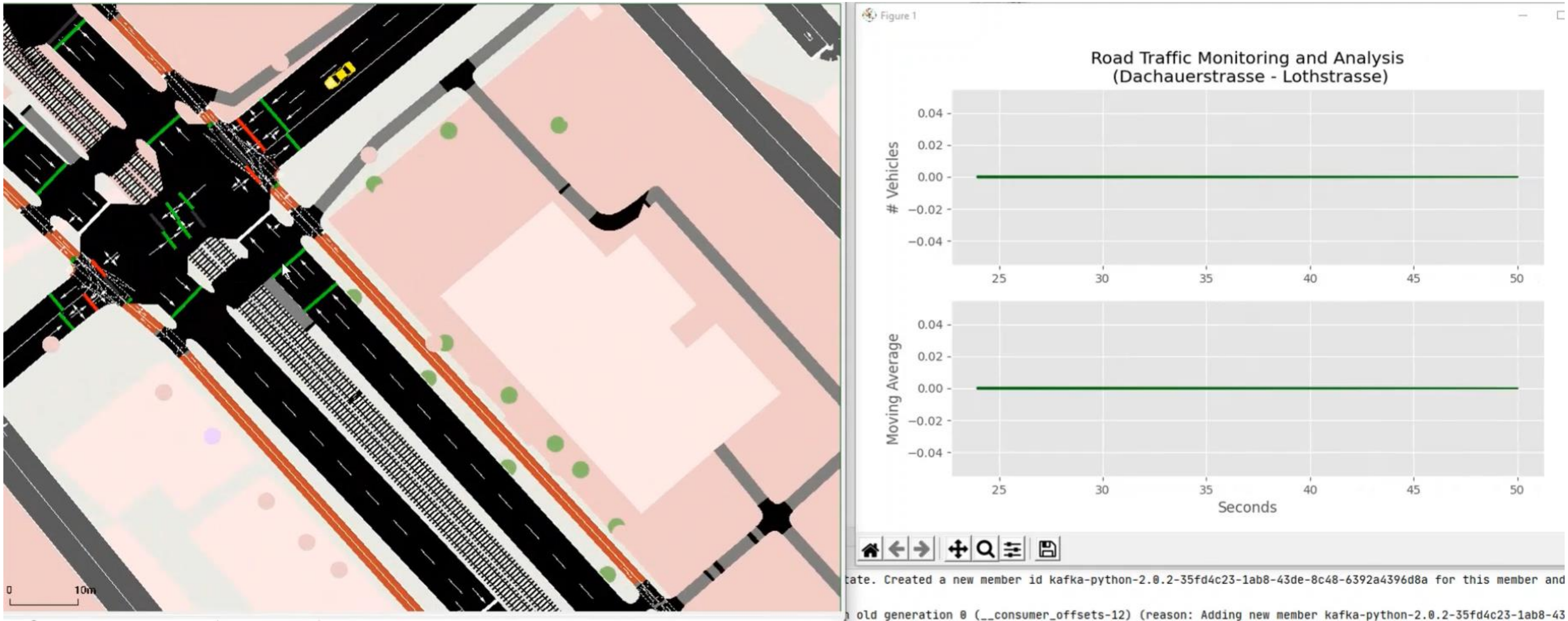
Road Traffic Monitoring and Analysis

Demo in Munich



Road Traffic Monitoring and Analysis

Demo in Munich

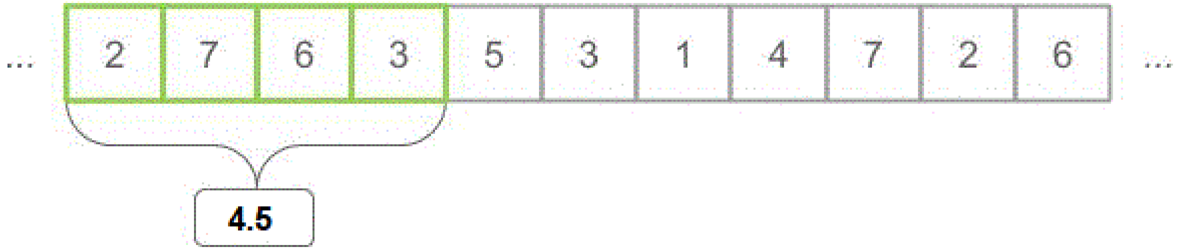


```
tate. Created a new member id kafka-python-2.0.2-35fd4c23-1ab8-43de-8c48-6392a4396d8a for this member and
 old generation 0 (__consumer_offsets-12) (reason: Adding new member kafka-python-2.0.2-35fd4c23-1ab8-43
```

Road Traffic Monitoring and Analysis

Applied concepts

Incremental computation



Principle

If we consider $x_1, x_2 - - - x_i$ the sensing data samples (# cars)

The sample mean is $\boxed{\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i}$

The incremental version

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \Leftrightarrow \bar{x}_n = \frac{1}{n}\left(x_n + \sum_{i=1}^{n-1} x_i\right) \Leftrightarrow \bar{x}_n = \frac{1}{n}\left(x_n + (n-1)\bar{x}_{n-1}\right)$$

$$\boxed{\bar{x}_n = \bar{x}_{n-1} + \frac{1}{n}\left(x_n - \bar{x}_{n-1}\right)}$$

Conclusions

Streaming Data Science

- Enables **real-time reaction to changes** in the observed system (i.e. technical, financial, biological etc.)

- Provides **tools for incrmental analysis.**

- It goes **beyond the traditional processing** aiming at low-latency and high-throughput data processing.

- An **emerging field of research** with high economical impact!

Bibliography

1. *Streaming Systems* by Tyler Akidau, Slava Chernyak, Leuven Lax, O‘Reilly 2018.

2. *Stream Processing with Apache Flink* by Fabian Hueske, Vasiliki Kalavri, O'Reilly Media, Inc. 2019.

3. *Machine Learning for Data Streams*: with Practical Examples in MOA by Albert Bifet, Ricard Gavaldà, Geoff Holmes, Bernhard Pfahringer, MIT Press 2018.

Code available on **Github**